

UNIVERSIDAD AUTÓNOMA DE MADRID

ESCUELA POLITÉCNICA SUPERIOR



# Máster en Bioinformática y Biología Computacional

## TRABAJO FIN DE MÁSTER

# PREDICCIÓN DE EFECTOS DE DROGAS EN CÁNCER MEDIANTE MACHINE LEARNING Y FLUX BALANCE ANALYSIS

Autor: Sofía Cerdá Muñoz

Tutor: Raúl Guantes Navacerrada

Febrero 2021



# PREDICCIÓN DE EFECTOS DE DROGAS EN CÁNCER MEDIANTE MACHINE LEARNING Y FLUX BALANCE ANALYSIS

Autor: Sofía Cerdá Muñoz  
Tutor: Raúl Guantes Navacerrada

Grupo de Biofísica y Biología de Sistemas  
Dpto. de Física de la Materia Condensada  
Universidad Autónoma de Madrid  
Febrero 2021



## Resumen

Cada vez se encuentran más evidencias sobre el hecho de que las células cancerígenas reprograman el metabolismo para su beneficio, con el fin de mejorar su proliferación. Por este motivo, están apareciendo nuevas vías terapéuticas que se basan en este principio, analizando en detalle el metabolismo tumoral, para así saber crear drogas enfocadas en este aspecto. Para ello, es necesario el análisis de una gran cantidad de datos, tanto de expresión genética global (*genome-wide*), como de metabolismo, lo que ha llevado al desarrollo de técnicas computacionales capaces de analizar tal carga de información. El *Machine Learning* se ha convertido en una de las principales herramientas utilizadas para este propósito. Sin embargo, la alta complejidad de sus algoritmos, aunque favorecen la eficacia y precisión de sus resultados, dificultan su interpretabilidad, lo cuál es esencial para conseguir una respuesta biológica. Una herramienta complementaria que ayudaría a la interpretabilidad en el ámbito del metabolismo, es la aplicación del *Flux Balance Analysis* (FBA), el cual permite simular modelos metabólicos en un entorno concreto de estudio. Este trabajo se centra en esta idea, combinando para ello la creación de flujos metabólicos simulados mediante FBA, con un algoritmo de *Machine Learning*, concretamente el algoritmo *MultiTask Elastic Net*, con el objetivo de analizar y predecir el efecto de drogas antitumorales sobre distintas líneas celulares cancerígenas.

## Palabras Clave

MultiTask Elastic Net; Flux Balance Analysis; Metabolismo; Cáncer

## **Abstract**

There is increasing evidence that cancer cells reprogram their metabolism to their advantage, in order to enhance their proliferation. For this reason, new therapeutic avenues based on this principle are appearing, analyzing tumor metabolism in detail, in order to create drugs focused on this aspect. This requires the analysis of a large amount of data, both genome-wide gene expression and metabolism, which has led to the development of computational techniques capable of analyzing such a load of information. Machine Learning has become one of the main tools used for this purpose. However, the high complexity of its algorithms, although favoring the efficiency and accuracy of its results, hinders its interpretability, which is essential to achieve a biological response. A complementary tool that would help interpretability in the field of metabolism is the application of Flux Balance Analysis (FBA), which allows the simulation of metabolic models in a specific study environment. This work focuses on this idea, combining the creation of simulated metabolic fluxes using FBA with a Machine Learning algorithm, specifically the MultiTask Elastic Net algorithm, with the aim of analyzing and predicting the effect of antitumor drugs on different cancer cell lines.

## **Key words**

MultiTask Elastic Net; Flux Balance Analysis; Metabolism; Cancer

# Agradecimientos

Quisiera agradecer a mi director, Raúl Guantes Navacerrada, el apoyo y el esfuerzo dedicado a este proyecto. A pesar de las dificultades y todas las cosas que he tenido que aprender en el desarrollo de este trabajo, he disfrutado del camino.

Gracias también a toda mi familia por todo el apoyo y por siempre estar ahí.





# Índice general

<b>Índice de Figuras</b>	<b>IX</b>
<b>Índice de Tablas</b>	<b>XI</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Objetivos . . . . .	2
<b>2. Métodos</b>	<b>3</b>
2.1. Bases de datos . . . . .	3
2.2. Flux Balance Analysis . . . . .	4
2.3. Multitask Elastic Net . . . . .	6
2.4. Análisis de coeficientes y pathways . . . . .	8
<b>3. Resultados y Discusión</b>	<b>9</b>
3.1. Coeficientes de metabolómica . . . . .	9
3.1.1. Cluster 1 . . . . .	9
3.1.2. Clusters 2 y 4 . . . . .	9
3.1.3. Cluster 3 . . . . .	10
3.1.4. Cluster 5 . . . . .	10
3.2. Coeficientes del FBA . . . . .	12
3.2.1. Función objetivo: NADPH . . . . .	12
3.2.2. Función objetivo: LDH . . . . .	14
<b>4. Conclusiones</b>	<b>17</b>
<b>Bibliografía</b>	<b>21</b>
<b>A. Material suplementario</b>	<b>23</b>



# Índice de Figuras

2.1. <b>Metabolismo simplificado de los principales pathways energéticos involucrados en cáncer.</b> Obtenido de Laura Pérez-Chirinos (TFM del Máster en Física de la Materia Condensada y de los Sistemas Biológicos, UAM, <i>Cancer metabolism study through Flux Balance Analysis, 2019</i> ). Abreviaturas: Glc[ex], extracellular glucose; Lac, lactate; AKG, %-ketoglutarate; Gln, glutamine. . . . .	6
2.2. <b>Elastic Net vs Multitask Elastic Net.</b> Modificado de Yuan et. al ( <i>Multitask learning improves prediction of cancer drug sensitivity, 2016</i> ). . . . .	8
3.1. <b>Clusters más destacados de metabolitos.</b> . . . . .	11
3.2. <b>Transportador de malato-aspartato.</b> Obtenido de <a href="https://en.wikipedia.org/wiki/Malate-aspartate_shuttle">https://en.wikipedia.org/wiki/Malate-aspartate_shuttle</a> . . . . .	13
3.3. <b>Coefficientes (en valor absoluto) de los flujos del FBA con función objetivo NADPH.</b> . . . . .	14
3.4. <b>Coefficientes (en valor absoluto) de los flujos del FBA con función objetivo LDH.</b> . . . . .	15
A.1. <b>Análisis de pathways del cluster 2.</b> . . . . .	23
A.2. <b>Análisis de pathways del cluster 4.</b> . . . . .	24
A.3. <b>Análisis de pathways del cluster 5.</b> . . . . .	25



# Índice de Tablas

2.1. <b>Drogas seleccionadas junto con sus mecanismos de acción.</b> Df: Antifols (impairs the function of folic acids). This inhibits production of DNA, RNA, and proteins.; A7: Alkylating at N-7 position of guanine; Db: DNA binder; A2: Alkylating at N-2 position of guanine; Ds: DNA synthesis inhibitor; Dr: Ribonucleotide reductase inhibitor; Tu: Tubulin affecting; TOP2: Topoisomerase 2 inhibitor; STK: Serine threonine kinase; TOP1: Topoisomerase 1 inhibitor; HDAC: Histone deacetylase. . . . .	3
2.2. <b>Reacciones con las fronteras modificadas</b> ( $\text{mmol } gDW^{-1}h^{-1}$ ). . . . .	5
3.1. <b>Reacciones más predictivas con la función objetivo NADPH.</b> . . . .	13



# 1

## Introducción

En los últimos años, numerosos estudios han acumulado una evidencia cada vez mayor de que el metabolismo alterado es una marca singular de cualquier tipo de célula cancerígena [1, 2]. Estas células presentan alteraciones metabólicas con las que pueden acceder tanto a fuentes de nutrientes convencionales como a otras menos habituales, y utilizarlas para incrementar su proliferación [3]. El metabolismo alterado no es una propiedad estática de las células cancerígenas, sino que se trata de una característica dinámica y plástica, es decir, que en condiciones de escasez o cambios de nutrientes, las células cancerígenas pueden volver a reprogramar su metabolismo para sobrevivir con los recursos de los que disponen [3].

El estudio detallado del metabolismo tumoral está empezando a abrir nuevas vías terapéuticas en cáncer [4]. La investigación de dianas metabólicas se está llevando a cabo mediante dos métodos que integran información a escala global. Por un lado, el desarrollo de modelos computacionales detallados del metabolismo y expresión genética globales (*genome-wide*), y la simulación del efecto que tendrían en ellos diferentes drogas [5, 6, 7]. Por otro lado, la correlación de datos de expresión y metabólicos de muestras de tumores con la respuesta clínica a tratamientos de los pacientes de los que proceden [8].

Debido a la gran cantidad de información genómica de la que se dispone a día de hoy, que además crece exponencialmente gracias al constante desarrollo de nuevas técnicas de secuenciación, el aprendizaje automático (*Machine Learning, ML*) se ha convertido en una de las herramientas más potentes para el análisis de estos datos y la obtención de respuestas biológicas. Sin embargo, aunque el uso de estos algoritmos permite el manejo de una gran cantidad de datos, presentan el inconveniente de que actúan como una “caja negra”, de manera que no se puede saber cuáles fueron las razones que llevaron a obtener un determinado resultado.

Por ello, es necesaria una estrategia complementaria que permita interpretar el modelo [9] de manera que se pueda reflejar el mecanismo biológico subyacente. Una de estas estrategias son los modelos basados en restricciones (*Constraint-Based Models*), y actualmente ya existen diversos trabajos que combinan estas técnicas con aprendizaje automático [10]. Esta combinación es precisamente en la que se centra este trabajo, con el objetivo de predecir el efecto de drogas en líneas celulares cancerígenas sobre una serie de condiciones metabólicas.

## 1.1. Objetivos

---

El objetivo de este trabajo es realizar una investigación preliminar sobre el papel que puede jugar la heterogeneidad metabólica y la reprogramación del metabolismo celular en la respuesta y resistencia de diferentes líneas celulares cancerígenas a distintas drogas, siempre de forma general, sin enfocar en ningún tipo de tumor en concreto. Para ello, se combinará el uso de modelos metabólicos globales de líneas celulares cancerígenas [11] con el de diferentes *screenings* experimentales que miden la respuesta de una población celular al tratamiento con diferentes tipos de drogas. Estos *screenings* se obtendrán de la base de datos NCI60 del *National Cancer Institute* de EEUU, con 60 líneas celulares cancerígenas [12].

Además, también se combinarán técnicas de ML que permitan identificar qué características del metabolismo tumoral pueden tener un valor predictivo en la respuesta a drogas [13]. Concretamente, se utilizará un modelo de *Multitask Elastic Net*, que permite realizar una regresión sobre múltiples *targets*.

Para llevar a cabo el análisis del metabolismo y el impacto sobre este al aplicar una determinada droga, se utilizará el método de *Flux Balance Analysis* (FBA) [14]. Este método, asume que todas las reacciones de la red metabólica se encuentran en equilibrio (balance de flujos). Además, también permite definir límites superior e inferior para los flujos de estas reacciones metabólicas, dando lugar a restricciones que sirven para delimitar el espacio de todos los flujos permitidos dentro del modelo, o estados posibles de la red metabólica. Una vez determinadas las restricciones y el espacio de estados posibles, se define una función objetivo relevante al problema estudiado (por ejemplo, producción de biomasa), de manera que de todo ese espacio de flujos posibles, solo se seleccione el óptimo que optimice dicha función. El FBA resulta de gran utilidad en el estudio de redes bioquímicas, así como en el análisis de fenotipos de organismos con diferentes perturbaciones ambientales y genéticas [15].

De este modo, los flujos obtenidos por el FBA constituirán la entrada del modelo de ML y los valores de GI50 (concentración de droga que causa una reducción del 50 % de la inhibición máxima de la proliferación celular), obtenidos en distintos experimentos para el conjunto de drogas, serán los *targets*. La salida del modelo indicará la relevancia de cada uno de los predictores (o *features*) utilizados sobre cada uno de los *targets*, es decir, mide cuantitativamente (positiva o negativamente) la influencia de cada una de las reacciones del modelo metabólico en la respuesta de cada una de las drogas de estudio tras aplicarlas sobre las líneas celulares de NCI60.

De manera complementaria, también se aplicará un *Multitask Elastic Net* utilizando datos de metabolómica sobre el mismo conjunto de drogas, para las mismas líneas celulares, con el objetivo de encontrar asociaciones entre pathways y cada una de las drogas de estudio, de forma que muestren cuales de ellos se comportan de forma similar para ciertas drogas y de distinta manera para otras.



# 2

## Métodos

### 2.1. Bases de datos

Los datos de *screening* de drogas se han obtenido de la base de datos NCI60 desarrollada por el *National Cancer Institute* de EEUU, mediante la herramienta *CellMiner* [16]. Estos datos están constituidos por valores de GI50 (normalizados logarítmicamente por  $\log_{10}$ ) de 53.279 drogas, de las cuales sólo se tuvieron en cuenta las 22.257 que pasaron el control de calidad. Cuanto más alto es uno de estos valores, mayor sensibilidad presenta una línea celular frente a una determinada droga.

Para este trabajo se utilizaron únicamente los datos de 26 drogas (agrupadas en 11 mecanismos de acción diferentes), que fueron seleccionadas bajo distintos criterios de limpieza y calidad: debían tener valores para todas las líneas celulares de NCI60, que hubieran sido aprobadas por la FDA o que estuvieran en la fase de ensayos clínicos, que tuvieran un nombre de droga no nulo y estandarizado y, finalmente, que al menos hubieran participado en 10 experimentos, de los cuales el 50 % hubieran superado el control de calidad (ver Tabla 2.1).

NSC	Drug name	Mechanism	NSC	Drug name	Mechanism
740	Methotrexate	Df	63878	Cytarabine	Ds
762	Nitrogen mustard	A7	82151	Daunorubicin	TOP2
3053	Actinomycin D	Db	119875	Cisplatin	A7
3088	Chlorambucil	A7	122819	Teniposide	TOP2
6396	Thiotepa	A7	125066	Bleomycin	Db
8806	Melphalan	A7	125973	Paclitaxel	Tu
9706	Triethylenemelamine	A7	169780	Dexrazoxane	TOP2
25154	Pipobroman	A7	280594	Triciribine Phosphate	STK
26980	Mitomycin	A2	301739	Mitoxantrone	TOP2
27640	Floxuridine	Ds	312887	Fludarabine	Ds
32065	Hydroxyurea	Dr	609699	Topotecan	TOP1
34462	Uracil mustard	A7	630176	Depsipeptide	HDAC
49842	Vinblastine	Tu	649890	Alvocidib	STK

**Tabla 2.1: Drogas seleccionadas junto con sus mecanismos de acción.** Df: Antifols (impairs the function of folic acids). This inhibits production of DNA, RNA, and proteins.; A7: Alkylating at N-7 position of guanine; Db: DNA binder; A2: Alkylating at N-2 position of guanine; Ds: DNA synthesis inhibitor; Dr: Ribonucleotide reductase inhibitor; Tu: Tubulin affecting; TOP2: Topoisomerase 2 inhibitor; STK: Serine threonine kinase; TOP1: Topoisomerase 1 inhibitor; HDAC: Histone deacetylase.

Por otro lado, también se han utilizado datos de metabolómica procedentes del *DTP Molecular Target Program* del *National Cancer Institute* ([https://dtp.cancer.gov/databases\\_tools/molecular\\_target](https://dtp.cancer.gov/databases_tools/molecular_target)) para 352 metabolitos diferentes y 58 de las 60 líneas celulares del NCI60. Sin embargo, para este estudio sólo se han utilizado 139 de estos metabolitos, descartando aquellos para los que no se conocía su identificador en ninguna de las siguientes bases de datos: HMDB, PubChem o KEGG.

## 2.2. Flux Balance Analysis

Cómo se ha mencionado anteriormente, el FBA es una aproximación matemática que analiza el flujo de metabolitos a través de una red metabólica [14]. Para ello, es necesario representar de alguna forma las reacciones y los metabolitos que constituyen la red, el entorno en el que se encuentran estos compuestos y unos objetivos a optimizar.

En primer lugar, se deben definir cada una de las reacciones de la red metabólica, indicando cuantitativamente la contribución de cada uno de los metabolitos, así como la dirección de los flujos, de manera que se pueden distinguir los reactivos de los productos. En la ecuación 2.1, se presentan 3 reacciones (en realidad 4, porque la primera es reversible) y observando la tercera, por ejemplo, se puede ver que por cada 4 unidades del metabolito *B* se producen 3 del metabolito *A* y una del *C*. El resto de reacciones se interpretan de forma similar.



El siguiente paso es llevar la representación anterior al ámbito matemático. De esta manera, para una red de  $n$  reacciones y  $m$  metabolitos, se crea una matriz estequiométrica  $S$  de  $m \times n$ , en la que las filas representan los metabolitos, las columnas las reacciones y cada uno de los elementos de la matriz son los coeficientes estequiométricos de cada metabolito en cada una de las reacciones. Estos coeficientes tienen valor negativo si el metabolito es consumido, positivo si se produce y si cero si no participa en la reacción, por lo que las reacciones de la ecuación 2.1 tendrían asociada la matriz estequiométrica de la ecuación 2.2. En la práctica, estas matrices suelen ser dispersas, ya que la mayoría de las reacciones sólo involucran a unos pocos metabolitos, por lo que hay muchos elementos nulos. Finalmente, el flujo que atraviesa todas las reacciones se representa por un vector  $v$  de tamaño  $n$  y las concentraciones de todos los metabolitos, por un vector  $x$  de longitud  $m$ .

$$S = \begin{pmatrix} -1 & 1 & 1 & 3 \\ -2 & 2 & -1 & -4 \\ 1 & -1 & -1 & 1 \end{pmatrix} \begin{matrix} \leftarrow A \\ \leftarrow B \\ \leftarrow C \end{matrix} \tag{2.2}$$

Adicionalmente, para poder aplicar el FBA, se imponen una serie de restricciones que se deben cumplir. La primera de ellas es el balance de flujos, que establece que la cantidad total que se consuma de cualquier compuesto debe ser igual a la cantidad total que se produzca en estado estable ( $dx/dt = 0$ ), de manera que se equilibran las entradas y las salidas de la reacción, y matemáticamente se traduce en el sistema de ecuaciones lineales  $Sv = 0$ . Otro tipo de restricciones imponen límites superiores e inferiores a cada reacción, lo que establece un rango de flujos permitido ( $l_i < v_i < u_i$ ), pero pueden añadirse otras limitaciones.

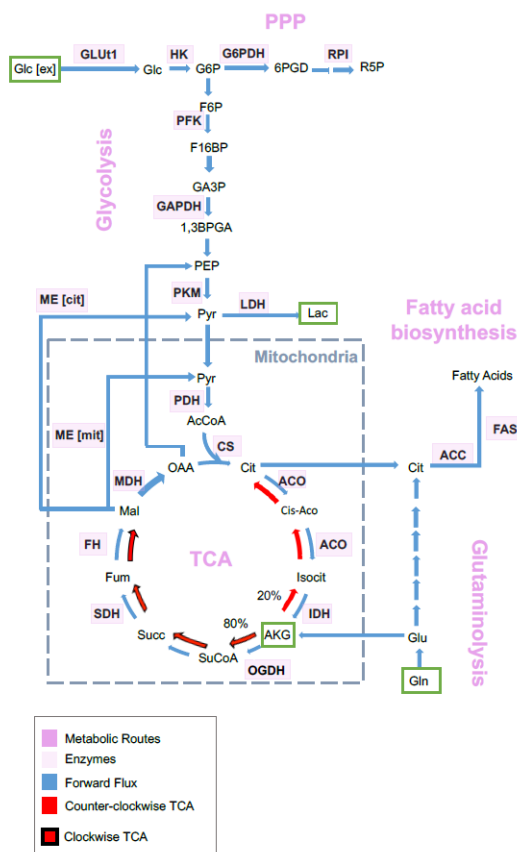
Sin embargo, en la práctica real, normalmente hay más reacciones que metabolitos ( $n > m$ ), es decir, hay más ecuaciones que incógnitas, por lo que quedaría un sistema de ecuaciones compatible indeterminado, lo que implica que existen infinitas soluciones. Por este motivo, el FBA elige una función objetivo a optimizar (ya sea minimizando o maximizando) definida como una combinación lineal de flujos, la cual se expresa como  $Z = c^T v$ , donde  $c$  es un vector de tamaño  $n$  cuyas entradas indican el peso de cada reacción, es decir, lo que contribuye cada una a la función objetivo. Por ejemplo, si la única reacción que interesa optimizar es la biomasa, el vector  $c$  tendría todas las entradas nulas salvo en la posición de la reacción de la biomasa, que tendría un 1, con lo que la función objetivo quedaría como  $Z = v_{biomasa}$ .

En FBA, la estrategia de optimización que se utiliza es la programación lineal, que consigue obtener, de todos los flujos  $v$  admisibles, es decir, que cumplan las restricciones (que se encuentren en el espacio nulo de  $S$ ), la solución al sistema de ecuaciones, o dicho de otro modo, el flujo óptimo que maximiza o minimiza la función objetivo. Existen distintas herramientas que implementan esta técnica, pero una de las más conocidas es *COBRA Toolbox* que actualmente tiene librerías de código abierto disponibles para Matlab [17, 18] y para Python [19] (la versión *CobraPy*). En este trabajo se ha utilizado la librería de Python, ya que este lenguaje también es el más potente para aplicaciones de ML.

Por último, la manera que tiene el FBA de simular condiciones diferentes en el entorno, es modificando los límites de las reacciones definidas como restricciones del modelo. En este estudio se ha ampliado el rango de flujos permitido para las reacciones más importantes de las principales rutas metabólicas que afectan a células cancerígenas: Glicólisis, Pentose Phosphate Pathway (PPP), producción de lactato, TriCarboxylic Acid cycle (TCA), o ciclo de Krebs, y Fatty Acid Biosynthesis (FAB) (ver figura 2.1y tabla 2.2). También se han utilizado como funciones objetivo las reacciones *Lactate dehydrogenase* (LDH\_L), para estudiar el reordenamiento metabólico al producir preferentemente lactato (metabolismo glicolítico), y la reacción *Mitochondrial NADPH demand* (DM\_nadph[m]), con el fin de estudiar el reordenamiento metabólico para producir biomasa y energía a través del ciclo TCA. El NADPH es la función objetivo establecida por defecto por Zielinski, D. C. et al. (2017), debido a que con ella obtuvieron la mejor concordancia con la cuantificación de metabolitos por etiquetado de  $^{13}C$  en el subconjunto de datos de las líneas celulares NCI60.

Reaction id	Reaction name	Lower bound	Upper bound	Pathway
GLUt1	glucose transporter 1	-2000	2000	Glycolysis
HK	hexokinase	0	2000	Glycolysis
PFK	phosphofructokinase	0	2000	Glycolysis
GAPDH	glyceraldehyde 3-phosphate dehydrogenase	0	2000	Glycolysis
PKM	pyruvate kinase	0	2000	Glycolysis
PDH	pyruvate dehydrogenase	0	2000	Glycolysis
G6PDH	glucose 6-phosphate dehydrogenase	0	2000	PPP
RPI	ribose-5-phosphate isomerase	-2000	2000	PPP
LDH	lactate dehydrogenase	-2000	0	Lactate
CS	citrate synthase	0	2000	TCA
ACO	aconitase	-2000	2000	TCA
IDH	isocitrate dehydrogenase	-2000	2000	TCA
OGDH	oxoglutarate dehydrogenase	0	2000	TCA
SDH	succinate dehydrogenase	0	2000	TCA
FH	fumarase	-2000	2000	TCA
MDH	malate dehydrogenase	-2000	2000	TCA
ME[cit]	cytosolic malic enzyme	0	2000	TCA
ME[mit]	mitochondrial malic enzyme	0	2000	TCA
ACC	acetyl-CoA carboxylase	0	2000	FAB
FAS	fatty-acyl-CoA synthase (n-C18:0CoA)	0	1000	FAB

Tabla 2.2: Reacciones con las fronteras modificadas ( $\text{mmol } gDW^{-1}h^{-1}$ ).



**Figura 2.1: Metabolismo simplificado de los principales pathways energéticos involucrados en cáncer.** Obtenido de Laura Pérez-Chirinos (TFM del Máster en Física de la Materia Condensada y de los Sistemas Biológicos, UAM, *Cancer metabolism study through Flux Balance Analysis*, 2019). Abreviaturas: Glc[ex], extracellular glucose; Lac, lactate; AKG, %-ketoglutarate; Gln, glutamine.

### 2.3. Multitask Elastic Net

El *Multitask Learning* (MTL), o aprendizaje multitarea, se basa en la resolución de varias tareas al mismo tiempo, aprendiendo tanto de las características comunes como de las diferencias entre ellas. Es decir, el modelo aprende en paralelo todas las tareas, pero también comparte la información aprendida con cada una de ellas para resolver las otras [20]. Esta técnica es especialmente útil cuando las tareas están relacionadas y, por lo general, están poco muestreadas, aunque también se ha demostrado que puede ser muy poderoso para tareas no relacionadas [21].

El MTL puede aplicarse tanto a problemas de clasificación como de regresión, utilizando diferentes tipos de algoritmos de ML. En este estudio, los problemas que se presentan son de regresión múltiple, dado que lo que se pretende es conocer el grado de contribución de cada una de las *features* o variables predictivas (metabolitos o reacciones) para alcanzar distintos *targets* (valores de GI50 de diferentes drogas). Para esta regresión lineal múltiple, se ha aplicado también una estrategia de regularización,  $R(\mathbf{w})$ , ya que esta técnica permite incorporar penalizaciones en el ajuste por mínimos cuadrados (ecuación 2.3), de manera que se pueda evitar el *overfitting*, reducir la varianza y el efecto de la correlación entre los predictores, así como minimizar la influencia de las variables menos relevantes, lo que ayuda a conseguir un modelo con mayor poder de generalización.

La estrategia de regularización elegida ha sido *Elastic Net*, que combina las ventajas de la norma  $\ell_1$ , que corresponde a la suma de valores absolutos ( $\|\mathbf{w}\|_1 = \sum_{i=1}^d |\mathbf{w}_i|$ ), con las de

la norma  $\ell_2$  o regularización de Tikhonov, que corresponde a la suma de cuadrados ( $\|\mathbf{w}\|_2^2 = \sum_{i=1}^d \mathbf{w}_i^2$ ), donde  $\mathbf{w}$  es un vector de tamaño  $d$ . La combinación de penalizaciones de sendas normas permite controlar la complejidad global del modelo. Por un lado, la norma  $\ell_1$  fuerza la aparición de coeficientes nulos para aquellas variables poco o nada influyentes, realizando así una selección de predictores, y por el otro, la norma  $\ell_2$  reduce, al mismo tiempo y de forma proporcional, la influencia de las variables altamente correlacionadas, sin llegar nunca a excluir ninguno de ellos, con lo que se consigue un equilibrio óptimo.

Así, el problema se reduce a minimizar la función que se muestra en la ecuaciones 2.3 y 2.4. La matriz  $\mathbf{X}$ , de tamaño  $n \times d$ , representa la matriz de características (o *feature matrix*) en la que  $n$  es el número de muestras (que en nuestro caso corresponden a las líneas celulares de NCI60) y  $d$  el número de variables predictivas (de metabolitos o reacciones, en este trabajo). El vector  $\mathbf{y}$  es un vector de tamaño  $n$  que representa el valor de cada uno de los *targets* correspondientes a cada muestra, mientras que el vector  $\mathbf{w}$ , de tamaño  $d$ , contiene los coeficientes correspondientes a cada predictor, que minimizan la función y que cumple que  $\mathbf{X}\mathbf{w} = \mathbf{y}$ . Por otro lado,  $\lambda \in \mathbb{R}$  y  $\alpha \in [0, 1]$  son los parámetros que controlan en qué grado y de que forma contribuye cada penalización, de manera que cuanto más cercano a 0 sea  $\alpha$ , mayor peso tendrá la penalización de la norma  $\ell_2$  (si es exactamente 0 sería equivalente a la estrategia *Ridge*) y cuanto más cercano a 1, la penalización predominante sería la de la norma  $\ell_1$  (si fuera exactamente 1 equivaldría a la estrategia *Lasso*). El parámetro  $\lambda$  aplica a ambas normas por igual, por lo que sirve para indicar cuánto debe afectar la penalización de forma global.

$$\min_{\mathbf{w} \in \mathbb{R}^d} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 + \gamma R(\mathbf{w}) \right\} \quad (2.3)$$

$$\begin{cases} R(\mathbf{w}) = \gamma_1 \|\mathbf{w}\|_1 + \frac{\gamma_2}{2} \|\mathbf{w}\|_2^2 \\ \gamma_1 = \lambda\alpha; \quad \gamma_2 = \lambda(1 - \alpha) \end{cases} \quad (2.4)$$

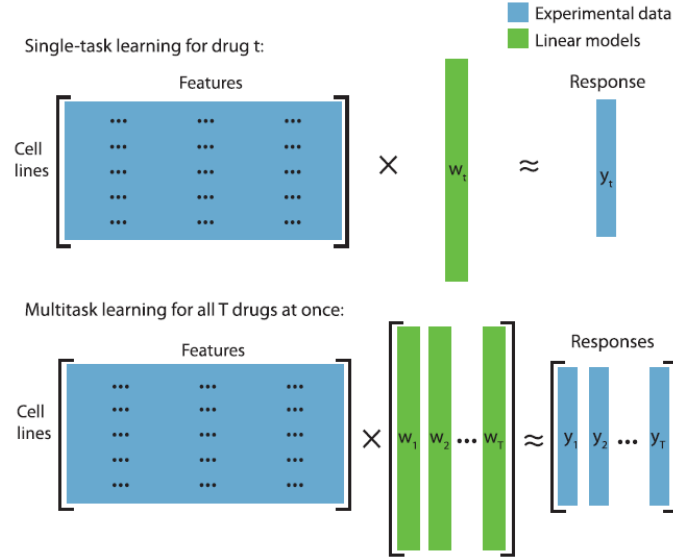
Sin embargo, extendiendo esta regularización a un problema de MTL, los *targets* estarían representados por una matriz  $\mathbf{Y}$  de tamaño  $n \times t$ , donde  $t$  es el número de tareas (número de drogas), y, por consiguiente,  $\mathbf{W}$  también pasaría a ser una matriz de coeficientes de tamaño  $d \times t$ , que correspondería al peso que se le da a cada variable predictiva para cada tarea (ver figura 2.2). Por este motivo, la versión MTL del *Elastic Net* es el *Multitask Elastic Net*, que utiliza la penalización de la norma  $\ell_{2,1}$ , que no es más que la norma  $\ell_1$  de la norma  $\ell_2$  de las diferentes tareas ( $\|\mathbf{W}\|_{2,1} = \sum_{j=1}^t \sqrt{\sum_{i=1}^d \mathbf{w}_{ij}^2}$ ), y la norma  $\ell_2$  matricial, también llamada "Norma de Frobenius" ( $\|\mathbf{W}\|_{Fro} = \sqrt{\sum_{i=1}^d \sum_{j=1}^t |\mathbf{w}_{ij}|^2}$ ). Con todo esto, la función objetivo a minimizar por la regularización *Multitask Elastic Net* se muestra en la ecuación 2.6.

$$\min_{\mathbf{W} \in \mathbb{R}^{d \times t}} \left\{ \frac{1}{2n} \|\mathbf{Y} - \mathbf{X}\mathbf{W}\|_{Fro}^2 + \gamma R(\mathbf{W}) \right\} \quad (2.5)$$

$$\begin{cases} R(\mathbf{W}) = \gamma_1 \|\mathbf{W}\|_{2,1} + \frac{\gamma_2}{2} \|\mathbf{W}\|_{Fro}^2 \\ \gamma_1 = \lambda\alpha; \quad \gamma_2 = \lambda(1 - \alpha) \end{cases} \quad (2.6)$$

En este trabajo, el software empleado para aplicar este modelo al problema que se propone, ha sido la librería *scikit-learn* [22] implementada en lenguaje Python. Concretamente, se han entrenado los datos con el modelo *MultiTaskElasticNetCV* ([https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.MultiTaskElasticNetCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.MultiTaskElasticNetCV.html)) que, además, permite hacer una búsqueda de los mejores parámetros  $\lambda$  y  $\alpha$  mediante validación cruzada. Esta técnica consiste en dividir el conjunto de datos en distintos grupos equitativamente (el número de grupos es un parámetro que se debe indicar al modelo), de forma que se entrena el modelo con todos los

grupos menos uno, que es el que se va a utilizar para calcular el error. Este proceso se repite hasta que todos los grupos han tomado el papel de conjunto de *test*, y, a su vez, en cada uno de estos entrenamientos los parámetros  $\lambda$  y  $\alpha$  varían según unos valores indicados. Finalmente, para cada par de valores  $\lambda$  y  $\alpha$  probados, se calcula la media de error cometido entre todas las particiones, y los valores elegidos para el modelo final son aquellos que presentan un menor error medio. En este caso, se ha hecho una búsqueda de parámetros con 5 grupos, 100 valores distintos de  $\lambda$  y para  $\alpha$  se han ido probando entre los siguientes valores: [0.1, 0.5, 0.7, 0.9, 0.95, 0.99, 1].



**Figura 2.2: Elastic Net vs Multitask Elastic Net.** Modificado de Yuan et. al (*Multitask learning improves prediction of cancer drug sensitivity*, 2016).

## 2.4. Análisis de coeficientes y pathways

Para llevar a cabo el análisis y visualización de los coeficientes obtenidos tras aplicar el *Multitask Elastic Net*, tanto a los datos de metabolómica como a los flujos resultantes de la simulación del FBA, se han utilizado distintas herramientas. En primer lugar, para la visualización de los heatmap se ha empleado la librería *seaborn* (<https://seaborn.pydata.org>), concretamente la función "*clustermap*" (<https://seaborn.pydata.org/generated/seaborn.clustermap.html>), que además permite mostrar los coeficientes clusterizados en ambos ejes del heatmap. Para la obtención de los clusters, se ha utilizado el *Método de Ward* (o método de Ward de varianza mínima) para la vinculación jerárquica, donde el criterio para la elección del par de clusters a vincular en cada paso, se basa en el valor óptimo de una función objetivo, el error de la suma de los cuadrados o varianza [23]. La métrica utilizada para el cálculo de la distancia entre cada par de clusters, ha sido la distancia euclídea.

Por otro lado, para el análisis de pathways de los clusters de metabolitos estudiados, se ha utilizado la plataforma *MetaboAnalyst* (<https://www.metaboanalyst.ca>). Concretamente, se ha empleado la función de *Enrichment Analysis* (MSEA), frente a la base de datos *KEGG* (<https://www.genome.jp/kegg/pathway.html>) con 84 conjuntos de metabolitos basados en pathways de humanos. Adicionalmente, también se ha utilizado la función de *Pathway Analysis*, que integra el análisis de enriquecimiento de rutas metabólicas con un análisis de topológico de estas, empleando como método estadístico el *Hipergeometric Test*, y como método de análisis topológico el *Relative-betweenness Centrality*, que mide la importancia de cada nodo por su influencia relativa entre los distintos caminos del grafo.

# 3

## Resultados y Discusión

### 3.1. Coeficientes de metabolómica

---

Para los datos de metabolómica se han estudiado los 5 clusters más destacados (ver Figura 3.1), analizando los pathways significativos ( $p\text{-value} < 0.05$ ) obtenidos en cada uno de ellos, así como las funciones que desempeñan en el metabolismo. El análisis de enriquecimiento de pathways se ha llevado a cabo con la ayuda de la plataforma *MetaboAnalyst*, y algunos de los resultados se pueden ver en el anexo A.

#### 3.1.1. Cluster 1

Se puede observar que en este grupo predominan los coeficientes negativos altos, lo que indica una alta correlación entre el efecto de la mayor parte de las drogas, con niveles bajos de estos metabolitos. En este caso no ha sido posible la asignación de pathways significativos, dado que se trata de un grupo pequeño de metabolitos que participan en diferentes pathways poco correlacionados. Sin embargo, lo que sí se puede destacar de este cluster, es que en él se encuentran varios lípidos tales como el colesterol, el glycerol y el ácido oleico. Estudios recientes demuestran la implicación del metabolismo alterado de lípidos en cáncer y en respuesta a drogas, mencionando, entre otras cosas, que el aumento de la síntesis o captación de lípidos contribuye al rápido crecimiento de las células cancerosas y a la formación de tumores [24, 25]. Por otro lado, también se observa que para la droga *TRICIRIBINE-PHOSPHATE*, con mecanismo de acción *STK*, cuya función es la inhibición de kinasas de serina y treonina, la mayoría de los metabolitos de este grupo presentan coeficientes positivos.

#### 3.1.2. Clusters 2 y 4

Un dato curioso que merecía la pena analizar, es el comportamiento opuesto que presentan los clusters 2 y 4. Se puede observar cómo claramente los metabolitos del cluster 2 tienen, en general, coeficientes positivos para todas las drogas, salvo para las 4 primeras (*Vinblastine*, *Paclitaxel*, *Actinomycin D* y *Depsipeptide*) correspondientes a los mecanismos de acción *Tu*, *Db* y *HDAC* (ver tabla 2.1). Sin embargo, el efecto en los metabolitos del cluster 4 es justa-

mente el contrario, presentando coeficientes negativos para todas las drogas excepto para las ya mencionadas.

Entre los pathways más relevantes del cluster 2 (fig. A.1) se encuentra el *Aminoacyl-tRNA biosynthesis* ( $p < 3.34E-04$ ) implicado en el proceso de traducción, lo que indica que estos metabolitos están involucrados en la expresión genética. Otro pathway relevante de este cluster es el *D-Glutamine and D-glutamate metabolism* ( $p < 8.4E-04$ ), implicado en el metabolismo de aminoácidos secundarios. La glutamina es un aminoácido muy importante para las células cancerígenas, ya que muchas de ellas no pueden sobrevivir en su ausencia, sufriendo lo que se denomina como "adicción a la glutamina" [26]. Es esencial en la respuesta al estrés, la energía celular, el metabolismo de los lípidos en las células cancerosas y el anabolismo. Por su parte, para el cluster 4 se han encontrado, como pathways más enriquecidos (fig. A.2), el *Arginine and proline metabolism* ( $p < 1.32E-04$ ) y el *Arginine biosynthesis* ( $p < 4.75E-03$ ), ambos implicados en el metabolismo de aminoácidos esenciales. Existen estudios que demuestran que la arginina favorece la proliferación celular, siendo limitante en diversos tipos de cáncer con células cancerosas con auxotrofia, los cuales necesitan este aminoácido para proliferar rápidamente [27].

El mecanismo *Tu* corresponde a los inhibidores de tubulina, cuya función es la de prevenir la mitosis y la división celular, por tanto tiene sentido que el cluster 2 se vea afectado negativamente, dado que, por consiguiente, esto también impide el proceso de traducción. El mecanismo *Db* implica la unión de la droga al ADN, causando daños y alteraciones en el mismo. Asimismo, el mecanismo *HDAC*, asociado a los inhibidores de la histona desacetilasa, que, aunque son drogas represoras transcripcionales, también actúan a través de otros mecanismos no transcripcionales, como el daño al ADN, fomentando la detención del crecimiento, la apoptosis o la muerte celular inducida por las especies reactivas del oxígeno (ROS). De nuevo, el efecto negativo de estos mecanismos en el cluster 2 están justificados, tanto por la represión transcripcional, como por el hecho de que la presencia de glutamina favorece en muchas ocasiones la proliferación de células tumorales, lo cual se evita dañando directamente al ADN de éstas.

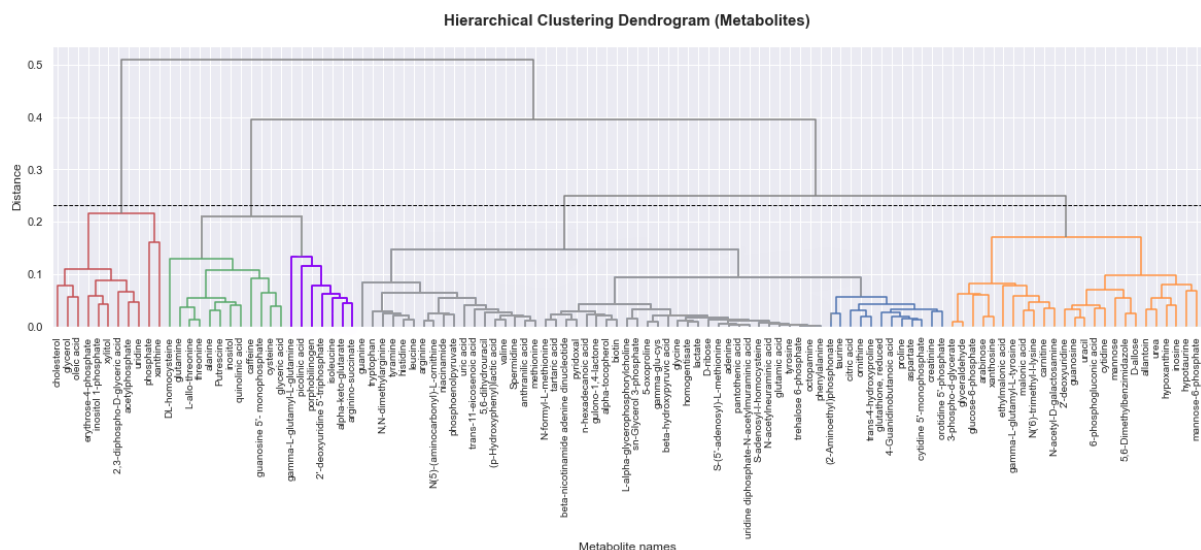
### 3.1.3. Cluster 3

Para este grupo de metabolitos, generalmente aparecen pesos positivos para todas las drogas, siendo el más destacado el *gamma-L-glutamyl-L-glutamine*. Este metabolito es un dipéptido intermediario en la degradación de la glutamina, muy importante en la proliferación de células cancerígenas, como se ha mencionado anteriormente, además de ser el precursor más inmediato del antioxidante glutatión. Otro metabolito interesante que aparece en este cluster, es el *alpha-keto-glutarato*, que también sale destacado en el análisis de coeficientes procedentes de los flujos obtenidos del FBA (ver sección 3.2). Este cetoácido se produce por la desaminación del glutamato y es un intermediario clave en el ciclo de Krebs, habiéndose demostrado que si la fuente de carbono proviene de la glutamina, el TCA funciona de forma no cíclica, produciéndose en este metabolito una bifurcación del flujo (ver figura 2.1) [28].

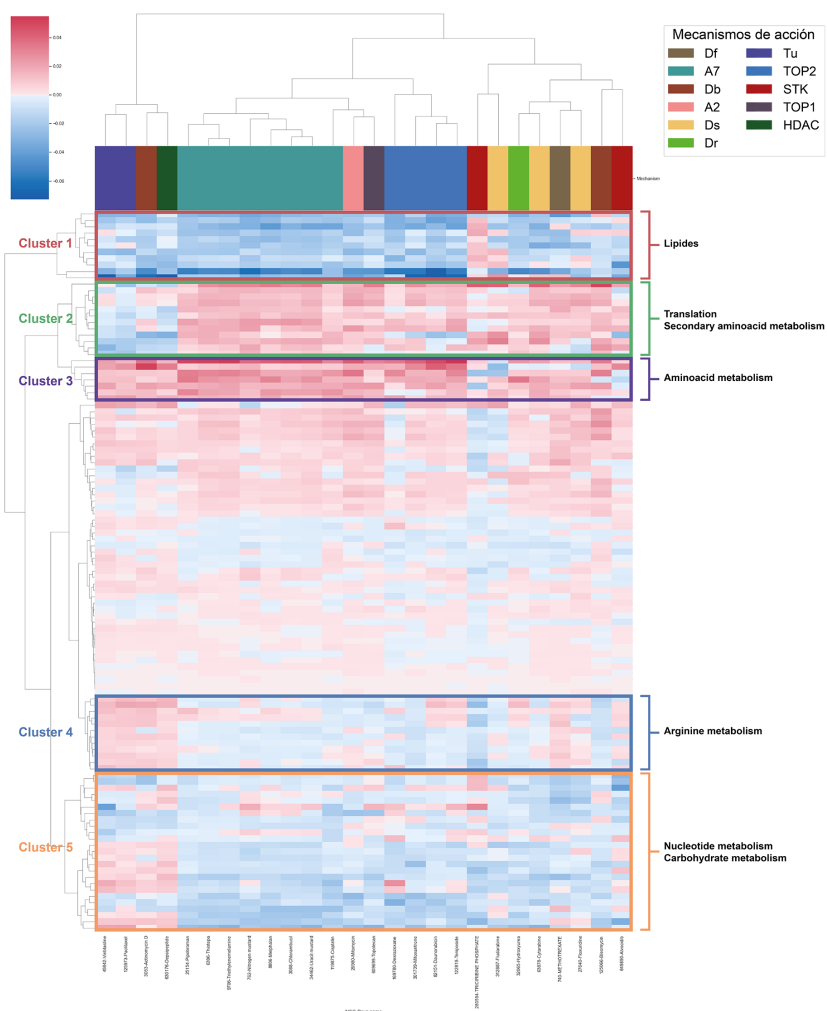
### 3.1.4. Cluster 5

En este conjunto de metabolitos, se puede observar cómo predominan los pesos negativos. Tras realizar el análisis de pathways (fig. A.3), se observó que este cluster estaba especialmente enriquecido por pathways involucrados en el metabolismo de nucleótidos, como el *Purine metabolism* ( $p < 4.5E-04$ ), así como el metabolismo de carbohidratos, como el *Fructose and mannose metabolism* ( $p < 3.71E-03$ ). Esto implica que la aplicación de estas drogas sobre líneas celulares cancerígenas, desfavorece a los pathways involucrados en producción de energía (catabolismo).





(a) Clusterización jerárquica de los metabolitos según sus coeficientes.



(b) Clustermap. Heatmap de los coeficientes, con cluster jerárquico de drogas y metabolitos.

**Figura 3.1: Clusters más destacados de metabolitos.**

## 3.2. Coeficientes del FBA

Para los resultados del FBA, también se ha llevado a cabo un análisis de las reacciones más predictivas del efecto de drogas según los coeficientes obtenidos tras aplicar el *Multitask Elastic Net* a los flujos finales de la simulación, distinguiendo si la función objetivo era el NADPH (*Mitochondrial NADPH demand*) o el LDH (*Lactate dehydrogenase*). En este caso, se han interpretado los pesos por su valor absoluto, dado que en los modelos metabólicos aparecían las reacciones reversibles dos veces, distinguiendo entre directa e inversa, lo que llevaba a obtener coeficientes de distinto signo para cada una de ellas, lo cuál impedía interpretar correctamente la influencia de dicho flujo en la simulación de efectos de drogas.

### 3.2.1. Función objetivo: NADPH

Las reacciones más destacadas cuando la función objetivo simulada es el NADPH, se muestran en la tabla 3.1, y corresponden con algunos de los clusters señalados en la figura 3.3. En primer lugar, el cluster 1 se caracteriza por presentar, fundamentalmente, reacciones transportadoras de metabolitos reductores, compuestos que donan un electrón a un agente oxidante en una reacción *REDOX*, a la mitocondria. Estas reacciones están involucradas especialmente en la cadena respiratoria (ciclo de Krebs) para la obtención de energía mediante el catabolismo de aminoácidos y nutrientes. La reacción ASPGLUm corresponde al transportador mitocondrial de aspartato-glutamato, el cual está involucrado en el pathway de transporte de malato-aspartato, trasladando grupos amino, así como el transportador alfa-cetoglutarato-malato o 2-oxoglutarato-malato (AKGMALtm), moviendo electrones. La transaminasa de aspartato (ASPTA) también forma parte del sistema de transporte de malato-aspartato, cuya función es eliminar un grupo amino del aspartato que se convierte en oxaloacetato, el cual luego se reduce a malato, mediante una enzima malato deshidrogenasa que utiliza un cofactor NADH, que puede atravesar la membrana mitocondrial (ver figura 3.2).

El transportador de malato-aspartato, es un sistema bioquímico que transporta los electrones producidos durante la glucólisis, a través de la membrana interna semipermeable de la mitocondria, para la fosforilación oxidativa en eucariotas. Estos electrones entran en la cadena de transporte de electrones de la mitocondria mediante agentes reductores para generar ATP. Este sistema es imprescindible para el transporte de equivalentes reductores de NADH, agentes principales en la cadena de transporte de electrones, a través de la membrana interna de la mitocondria, ya que esta es impermeable al NADH. Estos equivalentes reductores son entonces transportados por el malato.

Por tanto, parece que la respuesta a la mayor parte de las drogas, cuando las células tratan de proliferar optimizando NADPH, pasa por dirigir el flujo de la glicólisis hacia el ciclo de Krebs para obtener tanto energía como biomasa [29, 30]. También hay experimentos que demuestran que la inhibición de la transaminación del transportador de malato-aspartato ralentiza la proliferación de células cancerosas, debido a que el metabolismo de la glucosa también se ralentiza [31].

Por otro lado, otra reacción destacada dentro del cluster 1, es el transporte de ornitina (ORNtiDF). La ornitina también tiene su propio sistema transportador para entrar dentro de la mitocondria. Forma parte del ciclo de la urea, en el que participa convirtiéndolo en amoníaco, que es altamente tóxico, en urea, que es fisiológicamente inocuo. El amoníaco puede proceder tanto del efecto tóxico de las drogas, como del catabolismo de aminoácidos (degradación de aminoácidos para obtener energía). Además, hay estudios que revelan que el ciclo de la urea está alterado en muchos tipos de cáncer [32] y también que las células tumorales pueden usar esta vía para el anabolismo (producción de biomasa) [33].

Reaction id	Reaction name	Fórmula
ASPLGLUm	aspartate-glutamate mitochondrial shuttle	$asp_L[m] + glu_L[c] + h[c] \Leftrightarrow asp_L[c] + glu_L[m] + h[m]$
AKGMALtm	alpha-ketoglutarate/malate transporter	$akg[m] + mal_L[c] \Leftrightarrow akg[c] + mal_L[m]$
ASPTA	aspartate transaminase	$akg[c] + asp_L[c] \Leftrightarrow glu_L[c] + oaa[c]$
ORNtiDF	ornithine transport via diffusion (extracellular to cytosol)	$orn[e] \Leftrightarrow orn[c]$
G5SDym	glutamate-5-semialdehyde dehydrogenase (m)	$glu5p[m] + h[m] + nadph[m] \Rightarrow glu5sa[m] + nadp[m] + pi[m]$
GLU5Km	glutamate 5-kinase (m)	$atp[m] + glu_L[m] \Rightarrow adp[m] + glu5p[m]$
G5SADrm	L-glutamate 5-semialdehyde dehydratase, reversible, mitochondrial	$glu5sa[m] \Leftrightarrow 1pyr5c[m] + h2o[m] + h[m]$
P5CDm	1-pyrroline-5-carboxylate dehydrogenase	$1pyr5c[m] + 2,0h2o[m] + nadp[m] \Rightarrow glu_L[m] + h[m] + nadph[m]$

Tabla 3.1: Reacciones más predictivas con la función objetivo NADPH.

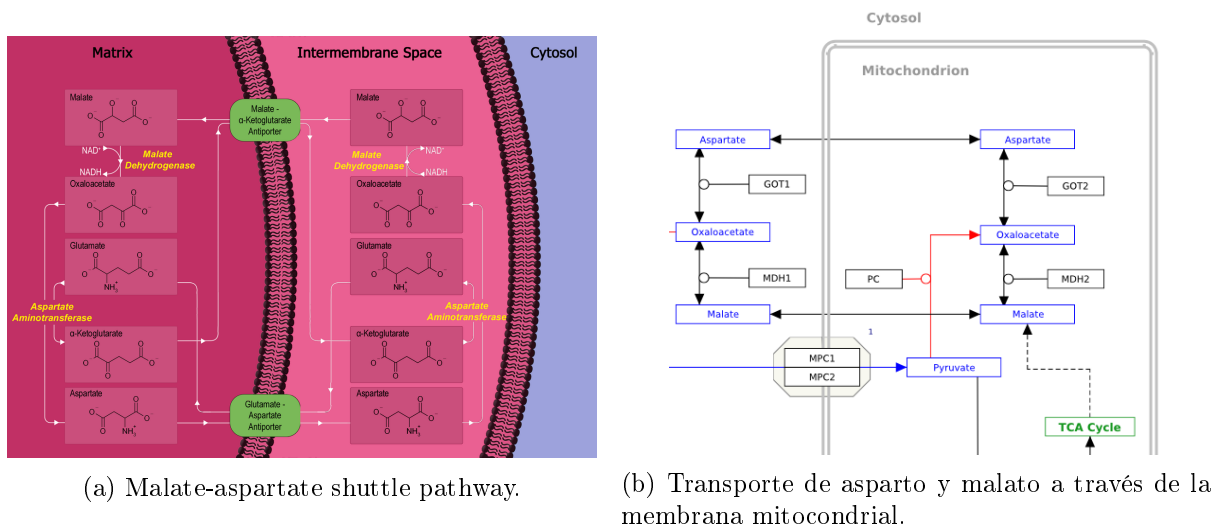
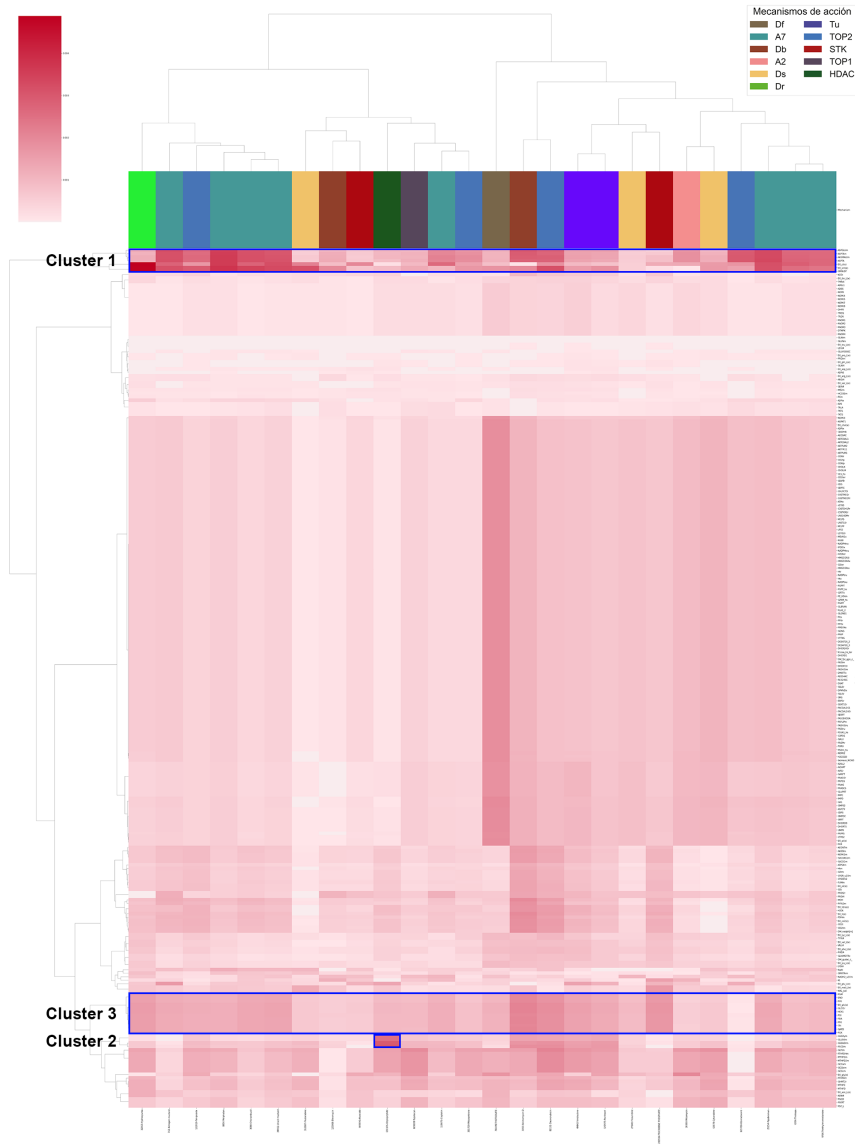


Figura 3.2: Transportador de malato-aspartato. Obtenido de [https://en.wikipedia.org/wiki/Malate-aspartate\\_shuttle](https://en.wikipedia.org/wiki/Malate-aspartate_shuttle)

En el cluster 2 se encuentran las reacciones G5SDym, GLU5Km y P5CDm, especialmente destacadas para el mecanismo de acción HDAC, asociado a los inhibidores de la histona de-sacetilasa, mencionado anteriormente. Todas ellas intervienen en el metabolismo de la arginina, que se produce a partir de glutamato también en el ciclo de la urea, de la cuál también se ha hablado en este trabajo al analizar el cluster 4 de los coeficientes de los metabolitos (ver sección 3.1.2), donde se mencionó su gran influencia en la proliferación celular de algunos tipos de células cancerígenas.

En respuesta a drogas, las células pueden reprogramar el metabolismo hacia la glicólisis para la obtención de energía y biomasa. Curiosamente, en el cluster 3 aparecen varias reacciones relacionadas con la glicólisis como destacadas, lo cuál se podría deber al hecho de haber ampliado las fronteras precisamente para ese pathway (ver tabla 2.2). Sin embargo, parece que es un efecto real, dado que las simulaciones cuya función objetivo era el LDH, con las mismas fronteras, no destacan esta ruta metabólica. A primera vista puede parecer extraño que la glicólisis se vea aumentada, cuando lo que se está optimizando es el NADPH, pero, cuando la glucosa entra en esta ruta, se genera piruvato, el cual, o bien produce lactato mediante glicólisis aeróbica, o se introduce en el TCA (ver figura 2.1), lo que explicaría su relevancia en este caso.



**Figura 3.3:** Coeficientes (en valor absoluto) de los flujos del FBA con función objetivo NADPH.

### 3.2.2. Función objetivo: LDH

En este caso, únicamente existe un cluster especialmente destacado (ver figura 3.4), que contiene reacciones de transferencia electrónica y transporte de FAD a la mitocondria, con el objetivo de conseguir energía directamente mediante el ciclo de Krebs. Existen evidencias que demuestran que las células cancerígenas pueden usar el lactato directamente como combustible para conseguir energía en el TCA [34, 35], y que no sólo se trata de un producto de desecho de la glucólisis anaeróbica y aeróbica. Esto es coherente con el hecho de que aparezcan destacadas reacciones que obtienen ATP mediante FAD/FADH<sub>2</sub> en la mitocondria, cuando se optimiza la reacción LDH.

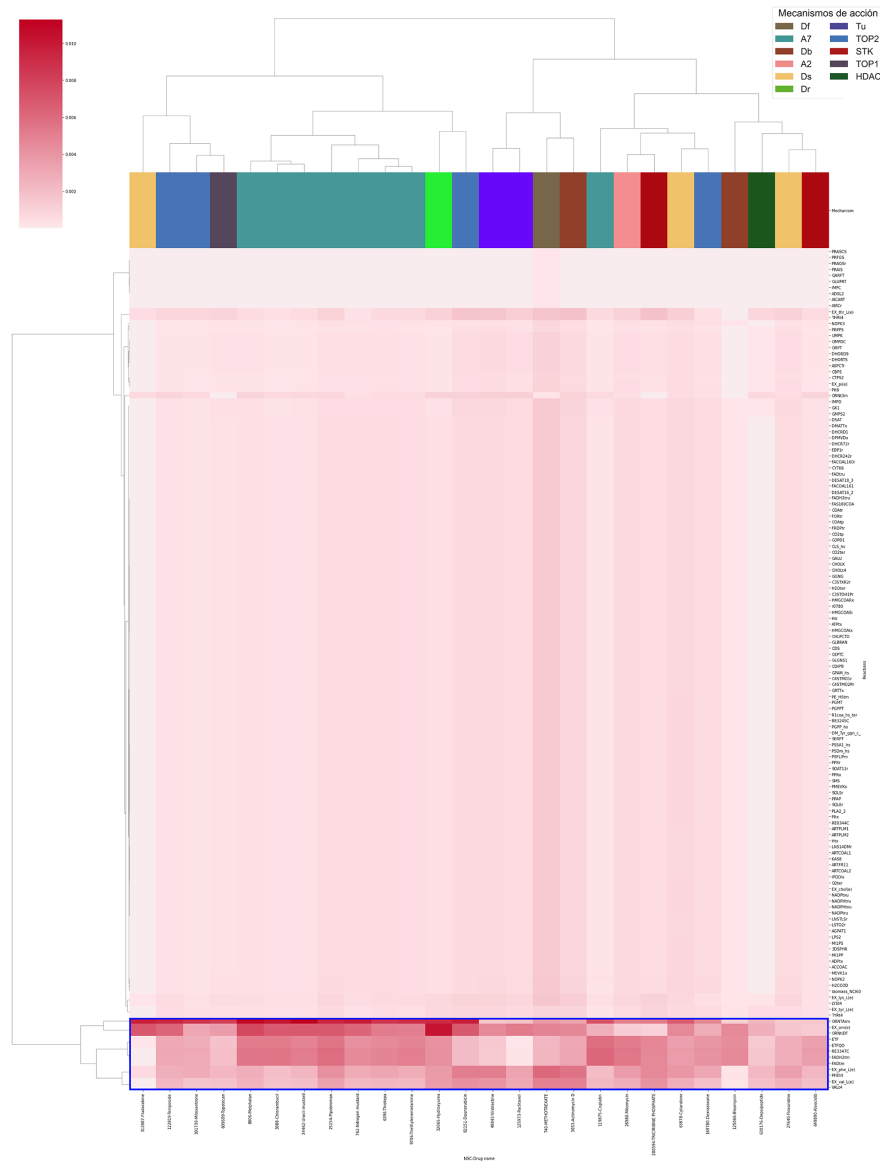


Figura 3.4: Coeficientes (en valor absoluto) de los flujos del FBA con función objetivo LDH.



# 4

## Conclusiones

Con este trabajo se ha mostrado que, combinando herramientas de *Machine Learning*, concretamente el algoritmo de *Multitask Elastic Net*, con simulaciones de modelos metabólicos realizadas mediante *Flux Balance Analysis*, es posible predecir el efecto de drogas frente a líneas celulares cancerígenas. Claro está, que aquí se ha llevado a cabo un estudio preliminar en calidad de prueba de concepto, cuyo objetivo era probar que esta combinación de métodos resultaba útil para el estudio de valores predictivos de efectos de drogas en el cáncer en general. Como trabajo futuro, sería interesante centrar el estudio en ciertas líneas celulares en particular, con fenotipos diferenciados, para las cuales se conozcan alteraciones metabólicas experimentales reales, y sobre ellas, aplicar el FBA en más detalle, construyendo si fuera preciso nuevos entornos (modificando fronteras o cambiando la función objetivo) en base a ese conocimiento, para poder validar el método.

En los experimentos de metabolómica, la clusterización jerárquica de sus coeficientes ha permitido llevar a cabo un análisis de enriquecimiento de pathways de los grupos más destacados, con lo que se han podido sacar conclusiones en base al nivel de correlación de la expresión de los metabolitos con el efecto de las drogas de estudio. También ha sido posible verificar que muchos de los pathways destacados estaban, de alguna manera, involucrados en el metabolismo de las células tumorales, lo cuál no ha hecho más que confirmar que el método es útil para revelar los mecanismos subyacentes que tienen las drogas para combatir la proliferación de las células tumorales.

Por su parte, los experimentos realizados con los flujos del FBA aplicando el *Multitask Elastic Net*, permiten observar el efecto de las drogas ante un metabolismo reprogramado, que es una de las claves de las células cancerígenas para su proliferación y supervivencia. En esta prueba de concepto, se han modificado las fronteras de algunas de las reacciones más importantes de los pathways más afectados por la reprogramación tumoral, para ampliar su espacio de acción. También se han probado distintas funciones objetivo, con el fin de ver los distintos efectos según cómo el modelo metabólico centre sus esfuerzos en producir un determinado elemento u otro. Efectivamente, los resultados de sendos modelos resultaron muy distintos, donde se pudo ver como el efecto de las drogas era más destacado, precisamente, en aquellas reacciones que fomentaban la producción de la función objetivo: NADH, para la reacción *Mitochondrial NADPH*, y el lactato, para la reacción *Lactate dehydrogenase*.





# Bibliografía

- [1] Douglas Hanahan and Robert A. Weinberg. Hallmarks of cancer: The next generation. *Cell*, 144(5):646–674, 2011.
- [2] Natalya N. Pavlova and Craig B. Thompson. The Emerging Hallmarks of Cancer Metabolism. *Cell Metabolism*, 23(1):27–47, 2016.
- [3] Lindsey K. Boroughs and Ralph J. Deberardinis. Metabolic pathways promoting cancer cell survival and growth. *Nature Cell Biology*, 17(4):351–359, 2015.
- [4] Alba Luengo, Dan Y. Gui, and Matthew G. Vander Heiden. Targeting Metabolism for Cancer Therapy. *Cell Chemical Biology*, 24(9):1161–1180, 2017.
- [5] Ori Folger, Livnat Jerby, Christian Frezza, Eyal Gottlieb, Eytan Ruppin, and Tomer Shlomi. Predicting selective drug targets in cancer through metabolic networks. *Molecular Systems Biology*, 7:501, 2011.
- [6] Itay Shaked, Matthew A. Oberhardt, Nir Atias, Roded Sharan, and Eytan Ruppin. Metabolic Network Prediction of Drug Side Effects. *Cell Systems*, 2(3):209–213, 2016.
- [7] Limin Li, Xiaobo Zhou, Wai Ki Ching, and Ping Wang. Predicting enzyme targets for cancer drugs by profiling human Metabolic reactions in NCI-60 cell lines. *BMC Bioinformatics*, 11:501, 2010.
- [8] Edoardo Gaude and Christian Frezza. Tissue-specific and convergent metabolic transformation of cancer correlates with metastatic potential and patient survival. *Nature Communications*, 7:13041, 2016.
- [9] Michael K. Yu, Jianzhu Ma, Jasmin Fisher, Jason F. Kreisberg, Benjamin J. Raphael, and Trey Ideker. Visible Machine Learning for Biomedicine. *Cell*, 173(7):1562–1565, 2018.
- [10] Guido Zampieri, Supreeta Vijayakumar, Elisabeth Yaneske, and Claudio Angione. Machine and deep learning meet genome-scale metabolic modeling. *PLoS Computational Biology*, 15:1–24, 2019.
- [11] Daniel C. Zielinski, Neema Jamshidi, Austin J. Corbett, Aarash Bordbar, Alex Thomas, and Bernhard O. Palsson. Systems biology analysis of drivers underlying hallmarks of cancer cell metabolism. *Scientific Reports*, 7:41241, 2017.
- [12] Robert H. Shoemaker. The NCI60 human tumour cell line anticancer drug screen. *Nature Reviews Cancer*, 6(10):813–823, 2006.
- [13] Jason H. Yang, Sarah N. Wright, Meagan Hamblin, Douglas McCloskey, Miguel A. Alcantar, Lars Schrübbers, Allison J. Lopatkin, Sangeeta Satish, Amir Nili, Bernhard O. Palsson, Graham C. Walker, and James J. Collins. A White-Box Machine Learning Approach for Revealing Antibiotic Mechanisms of Action. *Cell*, 177(6):1649–1661.e9, 2019.

- [14] Jeffrey D. Orth, Ines Thiele, and Bernhard O. Palsson. What is flux balance analysis? *Nature Biotechnology*, 28(3):245–248, 2010.
- [15] Edward J. O’Brien, Jonathan M. Monk, and Bernhard O. Palsson. Using genome-scale models to predict biological capabilities. *Cell*, 161(5):971–987, 2015.
- [16] William C. Reinhold, Margot Sunshine, Hongfang Liu, Sudhir Varma, Kurt W. Kohn, Joel Morris, James Doroshow, and Yves Pommier. CellMiner: A web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Research*, 72(14):3499–511, 2012.
- [17] Jan Schellenberger, Richard Que, Ronan M.T. Fleming, Ines Thiele, Jeffrey D. Orth, Adam M. Feist, Daniel C. Zielinski, Aarash Bordbar, Nathan E. Lewis, Sorena Rahmadian, Joseph Kang, Daniel R. Hyduke, and Bernhard Palsson. Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0. *Nature Protocols*, 6(9):1290–307, 2011.
- [18] Laurent Heirendt, Sylvain Arreckx, Thomas Pfau, Sebastián N. Mendoza, Anne Richelle, Almut Heinken, Hulda S. Haraldsdóttir, Jacek Wachowiak, Sarah M. Keating, Vanja Vlasov, Stefania Magnúsdóttir, Chiam Yu Ng, German Preciat, Alise Žagare, Siu H.J. Chan, Maike K. Aurich, Catherine M. Clancy, Jennifer Modamio, John T. Sauls, Alberto Noronha, Aarash Bordbar, Benjamin Cousins, Diana C. El Assal, Luis V. Valcarcel, Iñigo Apaolaza, Susan Ghaderi, Masoud Ahookhosh, Marouen Ben Guebila, Andrejs Kostromins, Nicolas Sompairac, Hoai M. Le, Ding Ma, Yuekai Sun, Lin Wang, James T. Yurkovich, Miguel A.P. Oliveira, Phan T. Vuong, Lemmer P. El Assal, Inna Kuperstein, Andrei Zinovyev, H. Scott Hinton, William A. Bryant, Francisco J. Aragón Artacho, Francisco J. Planes, Egils Stalidzans, Alejandro Maass, Santosh Vempala, Michael Hucka, Michael A. Saunders, Costas D. Maranas, Nathan E. Lewis, Thomas Sauter, Bernhard Palsson, Ines Thiele, and Ronan M.T. Fleming. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols*, 14(3):639–702, 2019.
- [19] Ali Ebrahim, Joshua A. Lerman, Bernhard O. Palsson, and Daniel R. Hyduke. COBRApy: COConstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*, 7(1):74, 2013.
- [20] Rich Caruana. Multitask Learning. *Machine Learning*, 28(1):41–75, 1997.
- [21] Bernardino Romera-Paredes, Andreas Argyriou, Nadia Bianchi-Berthouze, and Massimiliano Pontil. Exploiting unrelated tasks in multi-task learning. *Journal of Machine Learning Research*, 22:951–959, 2012.
- [22] G. Varoquaux, L. Buitinck, G. Louppe, O. Grisel, F. Pedregosa, and A. Mueller. Scikit-learn. *GetMobile: Mobile Computing and Communications*, 19(1):29–33, 2015.
- [23] Fionn Murtagh and Pierre Legendre. Ward’s Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward’s Criterion? *Journal of Classification*, 31(3):274–295, 2014.
- [24] Marteinn Thor Snaebjornsson, Sudha Janaki-Raman, and Almut Schulze. Greasing the Wheels of the Cancer Machine: The Role of Lipid Metabolism in Cancer. *Cell Metabolism*, 31(1):62–76, 2020.
- [25] Rimsha Munir, Jan Lisec, Johannes V. Swinnen, and Nousheen Zaidi. Lipid metabolism in cancer cells under metabolic stress. *British Journal of Cancer*, 120(12):1090–1098, 2019.

- [26] Emma R. Still and Mariia O. Yuneva. Hopefully devoted to Q: Targeting glutamine addiction in cancer. *British Journal of Cancer*, 116(11):1375–1381, 2017.
- [27] Songyun Zou, Xiangmei Wang, Po Liu, Changneng Ke, and Shi Xu. Arginine metabolism and deprivation in cancer therapy. *Biomedicine and Pharmacotherapy*, 118:109210, 2019.
- [28] Chiara Damiani, Riccardo Colombo, Daniela Gaglio, Fabrizia Mastroianni, Dario Pescini, Hans Victor Westerhoff, Giancarlo Mauri, Marco Vanoni, and Lilia Alberghina. A metabolic core model elucidates how enhanced utilization of glucose and glutamine, with enhanced glutamine-dependent lactate production, promotes cancer cell growth: The WarburQ effect. *PLoS Computational Biology*, 13(9):e1005758, 2017.
- [29] Magnus Monné, Angelo Vozza, Francesco Massimo Lasorsa, Vito Porcelli, and Ferdinando Palmieri. Mitochondrial carriers for aspartate, glutamate and other amino acids: A review. *International Journal of Molecular Sciences*, 20(18):4456, 2019.
- [30] Eric B. Taylor. Functional Properties of the Mitochondrial Carrier System. *Trends in Cell Biology*, 27(9):633–644, 2017.
- [31] Nina Ilic, Kivanç Birsoy, Andrew J. Aguirre, Nora Kory, Michael E. Pacold, Shambhavi Singh, Susan E. Moody, Joseph D. DeAngelo, Nicole A. Spardy, Elizaveta Freinkman, Barbara A. Weir, Aviad Tsherniak, Glenn S. Cowley, David E. Root, John M. Asara, Francisca Vazquez, Hans R. Widlund, David M. Sabatini, and William C. Hahn. PIK3CA mutant tumors depend on oxoglutarate dehydrogenase. *Proceedings of the National Academy of Sciences of the United States of America*, 114(17):E3434–E3443, 2017.
- [32] Joo Sang Lee, Lital Adler, Hiren Karathia, Narin Carmel, Shiran Rabinovich, Noam Auslander, Rom Keshet, Noa Stettner, Alon Silberman, Lilach Agemy, Daniel Helbling, Raya Eilam, Qin Sun, Alexander Brandis, Sergey Malitsky, Maxim Itkin, Hila Weiss, Sivan Pinto, Shelly Kalaora, Ronen Levy, Eilon Barnea, Arie Admon, David Dimmock, Noam Stern-Ginossar, Avigdor Scherz, Sandesh C.S. Nagamani, Miguel Unda, David M. Wilson, Ronit Elhasid, Arkaitz Carracedo, Yarden Samuels, Sridhar Hannenhalli, Eytan Ruppín, and Ayelet Erez. Urea Cycle Dysregulation Generates Clinically Relevant Genomic and Biochemical Signatures. *Cell*, 174(6):1559–1570, 2018.
- [33] Rom Keshet, Peter Szlosarek, Arkaitz Carracedo, and Ayelet Erez. Rewiring urea cycle metabolism in cancer to support anabolism. *Nature Reviews Cancer*, 18(10):634–645, 2018.
- [34] Inmaculada Martínez-Reyes and Navdeep S. Chandel. Waste Not, Want Not: Lactate Oxidation Fuels the TCA Cycle. *Cell Metabolism*, 26(6):803–804, 2017.
- [35] Joshua D. Rabinowitz and Sven Enerbäck. Lactate: the ugly duckling of energy metabolism. *Nature Metabolism*, 2(7):566–571, 2020.





## Material suplementario

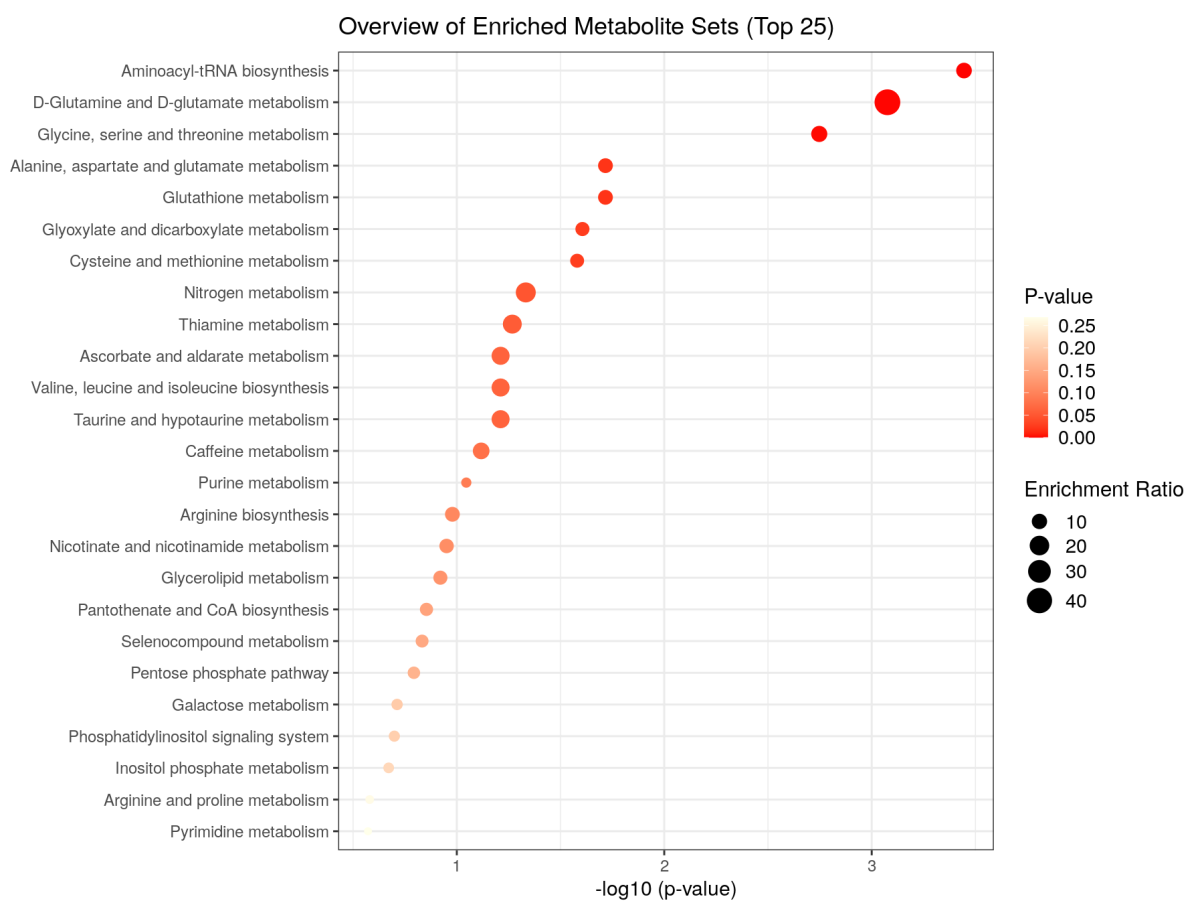


Figura A.1: Análisis de pathways del cluster 2.

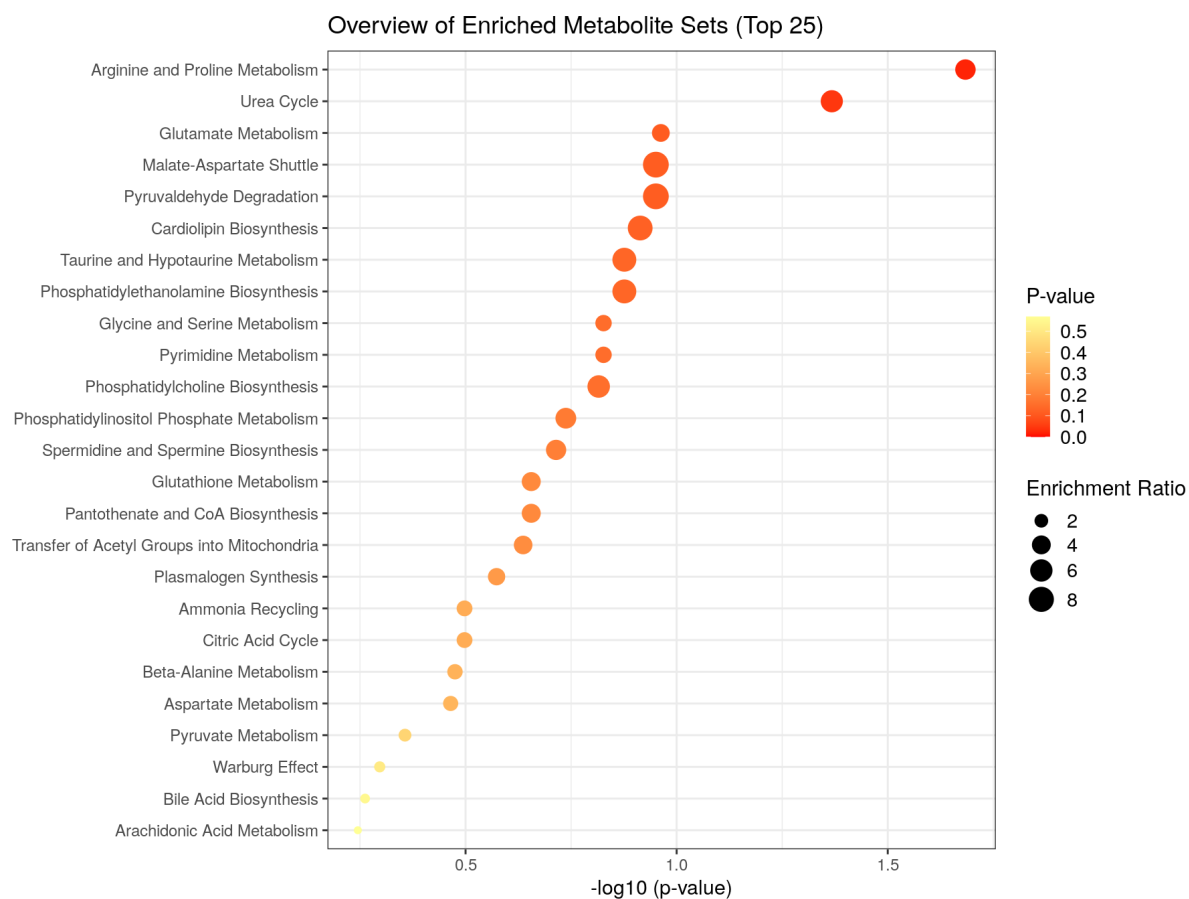


Figura A.2: Análisis de pathways del cluster 4.

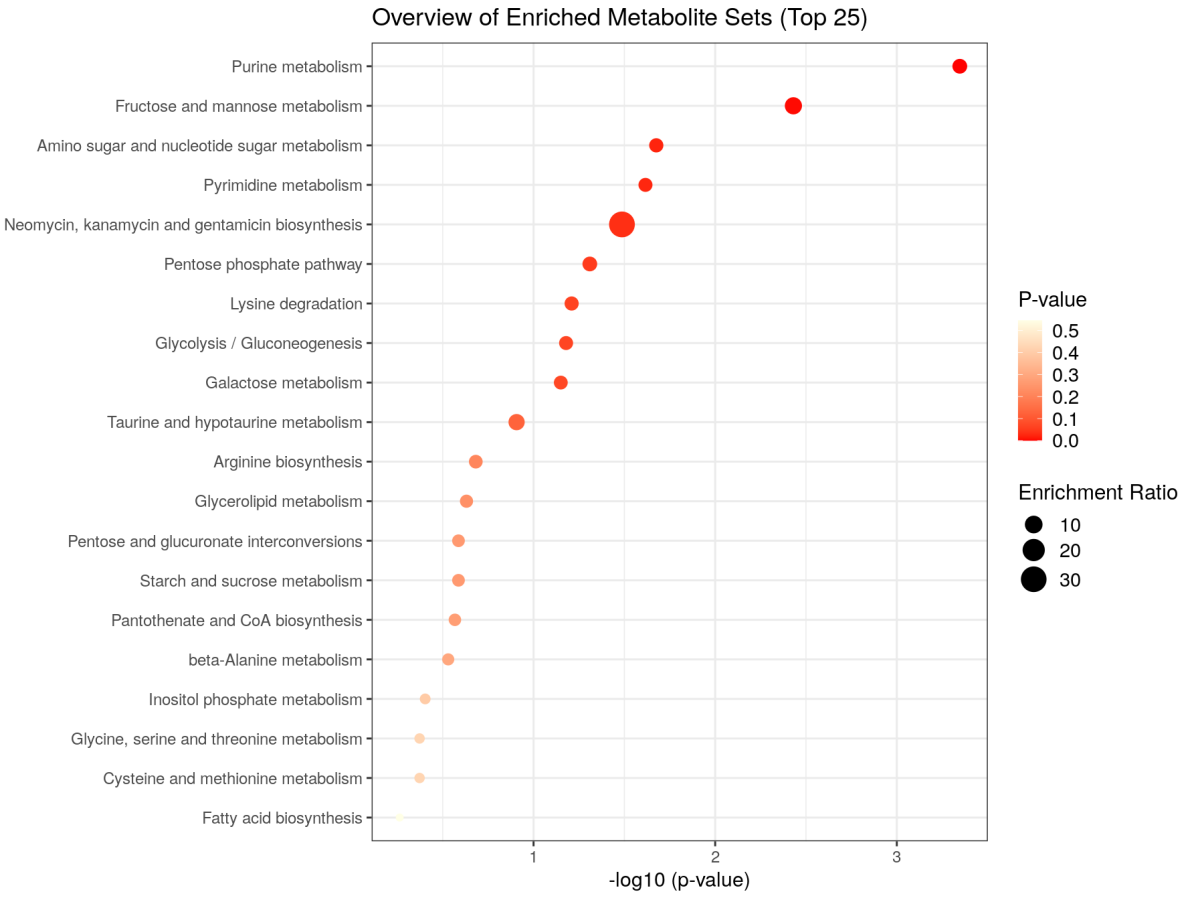


Figura A.3: Análisis de pathways del cluster 5.